(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 April 2004 (29.04.2004)

PCT

(10) International Publication Number
WO 2004/035771 A1

(51) International Patent Classification[7]: C12N 9/22, 15/55, C12Q 1/68

(21) International Application Number: PCT/EP2003/011210

(22) International Filing Date: 9 October 2003 (09.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
102 48 258.6 16 October 2002 (16.10.2002) DE

(71) Applicant *(for all designated States except US)*: **BIOPSYTEC ANALYTIK GMBH** [DE/DE]; Marie-Curie-Str. 1, 53359 Rheinbach (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BARON, Udo** [DE/DE]; Haydnstr. 57a, 53115 Bonn (DE). **IMHOFF, Ulrike** [DE/DE]; Georgstr. 33, 53111 Bonn (DE). **KOCH, Jürgen** [DE/DE]; Aegidienstr. 40, 53117 Bonn (DE). **LEUER, Marco** [DE/DE]; Kaiser-Karl-Ring 29, 53111 Bonn (DE). **WEBER, Jürgen** [DE/DE]; Schillerstrasse 16, 56567 Neuwied (DE). **OLEK, Klaus** [DE/DE]; Trierer Str. 90, 53115 Bonn (DE).

(74) Agents: **KRAUSS, Jan, B.** et al.; Pettenkoferstrasse 20-22, 80336 Munich (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 2004/035771 A1

(54) Title: MUTATED NUCLEIC ACID OF A CEL I-ENDONUCLEASE AND METHOD FOR PRODUCING THE RECOMBINANT, FULL-LENGTH CEL I-PROTEIN

(57) Abstract: The invention relates to a method for producing a recombinant, complete CEL I-protein, a plant endonuclease, and of parts thereof, by the expression of synthetic DNA-sequences. The invention also relates to the DNA-sequences themselves, which are produced for this purpose. Moreover, the invention relates to the use of the recombinantly produced CEL I-enzyme for detecting point mutations as well as larger mutated regions like e.g. deletions/insertions.

# Mutated Nucleic Acid of a Cel I-Endonuclease and Method for producing the Recombinant, Full-length Cel I-Protein

The invention relates to a method for producing a recombinant, complete CEL I-Protein, a plant endonuclease, or parts thereof, by the expression of synthetic DNA sequences. The invention also relates the DNA sequences themselves, which are produced for this purpose. Furthermore, the invention relates to the use of the recombinantly produced CEL I-enzyme for detecting point mutations as well as larger mutated regions like e.g. deletions/insertions.

The CEL I-enzyme is an endonuclease found in celery (Oleykowski et al. 1998), which recognises single "uneven elements" within the DNA-double helix and cleaves there in a specific manner. The enzyme therefore constitutes a very useful means for detecting mutations. It also specifically recognises single base mismatches (point mutations) and cleaves at one of the two DNA-strands proximately 3' from the mutation, thereby incising into the double strand.

For the application desired in this context, CEL I provides a number of advantages in comparison to other known nucleases also being able to incise into DNA-strands at uneven elements of the double helix structure:

The major problem of the presently known mismatch-recognising endonucleases, some of which are also commercially available, is based on their incomplete capability to identify all possible types of mismatches or mutations and on an unspecific DNA-degradation produced by them. As an example, the S1 nucleases do not cut at singles base mismatches (Loeb and Silber, 1981). The *Mung bean*-Nuclease provides an efficiency being five times higher at a pH of 5, than e.g. in the neutral pH-range (Kowalski and Sandford, 1982). The T4-Endonuclease VII does not only cut one strand of a double helix, but always cuts the complementary strand as well (Solaro et al., 1993). Furthermore the endonuclease isolated from T4-phages was shown to provide an unspecific activity of random DNA-degradation being significantly higher than that of the CEL I-enzyme, which is synthesised by the present

method (Cotton et al., 1999). Moreover, the degree of specificity of T4-endonuclease VII exhibits a high dependence on the length of the substrate and also on the sequence surroundings of the mismatches to be detected (Babon et al., 1999; Norberg et al., 2001).

This has to be regarded as a considerable disadvantage especially for a specific selection of mutations being unknown at first – a feature, that is of special relevance for the application described herein.

CEL I belongs to a distinct group of nucleases, which are found in many plant species and which are especially characterised by their specific maximum of activity at neutral pH value (Oleykowski et al., 1998), although an activity of CEL I is also to be found in a range of pH values between pH 5 and pH 9,5 (Oleykowski et al., 1998).

The capability to recognise and cut each form of base mismatches also irrespective of the AT-content in the vicinity of the mutation distinguishes the isolated CEL I-enzyme from other nucleases belonging to the family of plant nucleases and having their activity peak at a neutral pH value, like e.g. SP nuclease isolated from spinach (Yang et al., 2000; Oleykowski et al., 1999).

For reason of the toxic character of CEL I, it is until now impossible for cells without compartmentation to successfully express CEL I in a recombinant form, thus prohibiting to yield CEL I that way without the necessity to laboriously purify it in a native from the plant.

It is thus one objective of the present invention, to produce enzymatic material in the form of (the purest possible) active CEL I-enzyme in arbitrary amounts and by means of simple production methods, thereby providing sufficient amounts of this enzymatic material.

This objective of the invention is achieved by a special modification of the common DNA sequence of Cel I, this modification being especially designed for this aim. A further aspect of the present invention is thus this newly designed sequence.

A further aspect of the present invention refers to a method for producing the recombinant, complete CEL I-protein comprising the following steps: At first, a scheme of the DNA-sequence to be synthesised is created. This scheme is based on the cDNA sequence of the

CEL I-protein isolated from the celery plant *Apium graveolens* L. (see Olekowski et al., 2000).

For an expression in yeast, this DNA sequence is redrafted in dependence of the common yeast codon frequency (codon usage) while retaining the amino acid sequence of the CEL I-protein (FIG. 1). By the teaching of the present invention, an expression of Cel I in the yeast *Pichia pastoris* was thus allowed or favoured for the first time. A method for optimising codon usage is e.g. described by Outchkourov NS et al. in the case of the protein equistatin of the water anemone. This document neither reveals nor suggests the use of this method for non expressible enzymes with toxic effects on the cell.

Concretely, the invention thus relates to a method for producing a nucleic acid sequence, which codes for the complete CEL I-protein and allows to be recombinantly expressed in host cells, whereat the method comprises the following steps: providing the sequence coding for the CEL I-protein from a suitable organism, in particular from *Apium graveolens* L., and adequately modifying the codon frequency of the sequence to be expressed in comparison to the native sequence, whereat this modification is performed with regard to the host organism to be used for expression. This modification is accomplished by the following steps: a) partition of the planned sequence into an even number x, in particular 8, overlapping regions, b) synthesis of 2x mutated oligonucleotides, in particular oligonucleotides 1-16 to 16-16, each of which comprises the entire length of one overlapping region of both strands of the coding sequence, c) first PCR-amplification in order to produce x/2, in particular 4, overlapping fragments under employment of the oligonucleotides of step b), d) second PCR-amplification in order to produce x/4, in particular 2, overlapping regions under employment of the fragments of step c), and, e) third PCR-amplification in order to produce x/8, in particular 1, fragment, which comprises the coding region of CEL I. A schematic overview of a preferred embodiment is depicted in FIG. 3.

As an alternative, a method according to the invention may also be characterised in that instead of step e) the following steps are performed: e') Cloning the fragments generated in step d) into a suitable vector, and f) appropriately digesting the vectors and ligating the fragments in order to fuse them to form a complete "fragment" comprising the coding region of CEL I. Thereby, it is possible to perform further modifications of the sequence harboured in vector systems, if desired. Respective methods for this aim are familiar to the expert in the

field of nucleic acid applications and are described – among others – in the standard works "Molecular Cloning - A Laboratory Manual, Sambrook & Russel, 3rd Edition, (2001), Cold Spring Harbor Laboratory Press, or "Current Protocols in Molecular Biology", Ausubel et al. (1994 pp), Harvard Medical School.

Preferred is a method according to the invention, in which the codon frequency of the sequence to be expressed is modified according to the codon frequency of yeast, since yeast offers special advantages as a host organism for expression (see below).

Particularly preferred is a method according to the invention, which is characterised by the further attachment of nucleotides coding for present/additional N-terminal or C-terminal amino acid-tags, in particular for tags being comprised of 6 histidines. If necessary or useful, a base sequence in the form of a "His tag" can be added C-terminally or, in another case, N-terminally to the original sequence, this sequence coding for 6 histidines. Due to their strong affinity to nickel ions these tags are intended to support the purification of the expressed protein by means of immobilised nickel molecules, if desired. Paula de Mattos Areas et al., e.g. describe – among other things – the use of such "His-tags" in *Escherichia coli*. The fragment, being provided with a His-tag sequence at its N-terminus, has a cleavage side for factor Xa 3' from the tag sequence. This allows to subsequently process the expressed enzyme by cutting off the His-tag sequence being potentially obstructive for enzyme activity. The sequence designed such, will be designated in the following as 6His-Xa-Cel I sequence. The fragment being C-terminally linked to the His-tag will be designated as Cel I-6His in the following description of the present invention.

In order to facilitate the several steps of the method, the oligonucleotides synthesised in step (c) have an average length of 70 nucleotides and overlap in each case at about 20 bases. These values however, have to be understood as mere clues and may vary in dependence of the intended use. Suitable variants are easy to find for the expert on the basis of reactive or kinetic parameters; they are particularly dependent on the temperature and the base sequence.

A further aspect of the invention relates to a method for producing a recombinant, complete CEL I-protein from *Apium graveolens* L., comprising a) performing the above described method and b) expressing the nucleic acid sequence by means of a suitable expression system.

Especially preferred for this aim are vector-based expression systems being selected from the pPIC 9, pPIC 3, 5 and pQE-vectors. However, one may also employ other expression systems, which are familiar and depending on the host strain used.

A preferred option thereby is the expression of nucleic acid sequences in a host cell, which is selected from *Hansenula polymorpha*, *Pichia pastoris*, *Saccharomyces cerevisiae*, HeLa-cells, CHO-cells, *Toxoplasma gondii* and *Leishmania*. Particularly preferred is a method, in which the employed *Pichia pastoris* strain is the stain GS115.

The invention also relates to the complete DNA sequence of the CEL I-protein or expressible parts thereof derived from *Apium graveolens*, whereat said sequence is adapted for expression and is provided by a method according to the present invention. The wording "expressible parts" thereof refers to parts of the nucleic acid sequence coding for a polypeptide chain having an enzymatic function, in particular the function of the native CEL I-enzyme. Also comprised by the scope of the invention however, are nucleic acids coding for epitopes.

A preferred form is represented by a DNA sequence according to the invention, which codes for the *Apium graveolens* CEL I-protein, this sequence being characterised in that furthermore nucleotides are added, which encode additional N-terminal or C-terminal amino acid-tags, in particular tags being comprised of 6 histidines. In an also preferred form, the sequence at its both ends is equipped with restriction endonuclease cleavage sites, which are absent in the remaining sequence and in a vector to be employed. A most preferred *Apium graveolens* CEL I-protein encoding sequence according to the invention or parts thereof is presented in SEQ ID No.7. "Parts" in this context especially mean the fragments serving as probes, but also the overlapping oligonucleotides used for generating and cloning (see FIG. 2).

Both CEL I coding sequence variants were equipped at their ends with sequences of restriction sites, which are helpful for subsequent cloning steps. Examples for these sequences are the *Eco*RI restriction site positioned at both ends or the *Xho*I restriction site at the C-terminus (FIG. 2). Moreover, the translational start ATG was integrated in a Kozak sequence (consensus sequence for translational initiation) (ACC ATG G) (Kozak 1987; Kozak 1990) in both sequences. In the further procedure, 16 deoxyoligonucleotides were synthesised, which correspond to the planned sequence and completely cover the whole length of the respective cDNA. The deoxyoligonucleotides were synthesised such, that their sequences were

alternatively corresponding to the 5'-3'- or to the complementary 3'-5'-DNA-strand. The length of the deoxyoligonucleotides was between 40 to 93 bases with overlaps of an average of 20 bases between neighbouring sequences (FIG. 2). The artificial CEL I-gene was synthesised in the form of two independent partial fragments, the N-terminal and the C-terminal fragment, and fused afterwards via a *Hind*III restriction site (FIG. 2).

The generation of a partial fragment was accomplished according to the following principle: In a first step, four DNA sequences having the double length of two neighbouring deoxyoligonucleotides (minus the overlapping sequences) were generated for each fragment by means of asymmetric PCR. The amplification was accomplished such, that the neighbouring, accumulated DNA-strands in each case represent the opposite strand.

In a second amplification step, secondary products having a length of four original oligonucleotides (minus the overlapping sequences) can be synthesised (Fragments E and F, FIG. 3).

Another conversion of the such produced products into preparations mainly consisting of single-stranded DNA during an asymmetric PCR reaction with terminal oligonucleotides allowed for producing a double-stranded DNA-fragment of about 400 bp (fragment G; FIG.3) by means of two further PCR-amplifications.

The two such synthesised fragments were subsequently cloned in *E. coli* and their sequences determined. Erroneous sequence sections were excised with appropriate restriction enzymes and replaced by the corresponding correct sequence part of another clone. After having generated the N-terminal and the C-terminal fragment of the CEL I coding DNA-region in a correct form, both partial fragments were fused via the above cited *Hind*III restriction site in a suitable vector and cloned in *E. coli*.

Preferably, the artificial CEL I-gene generated in such a way can be transferred into suitable expression vectors during subsequent steps of the procedure. Favourable examples for these vectors are - among others - expression vectors suitable for the *Pichia pastoris* expression system like the pPIC 9 or the pPIC 3, 5 vector (Invitrogen). Also possible however, are other expression vectors and host organisms other than yeast, which are familiar to the expert. The subject of the invention is thus not restricted to a special host system.

A further aspect of the present invention thus relates to a host organism, which is capable to integrate and express a DNA sequence according to the invention. This host preferably is selected from *Hansenula polymorpha*, *Pichia pastoris*, *Saccharomyces cerevisiae*, HeLa-cells, CHO-cells, *Toxoplasma gondii* and *Leishmania*. Especially the *Pichia pastoris* stem GS115 is used. However, plant cells or insect cells may also be employed.

A preferred host organism in this invention, which is employed for expression, is the yeast *Pichia pastoris* (Invitrogen), whereof the preferred yeast strain is GS115 (Invitrogen). Yeast in general is preferred as the expression system, since it has - as a eukaryotic organism - many advantages compared to bacterial systems for expression, like e.g. the post-translational processing of proteins. Another important advantage of using a eukaryotic expression system is based on the cellular compartmentation being present in eukaryotic organisms. Expressing nucleases by means of recombinant expression systems in prokaryotic host organisms like bacteria is toxic for the cells due to the nucleases' DNA-degrading properties and has consequently and several times been described as being extremely difficult (Golz et a., 1995; Kosak and Kemper, 1990).

*Pichia pastoris* moreover is able to metabolise methanol as a hydrocarbon source. Thereby, the first step of methanol catabolism is catalysed by alcohol oxidase. *Pichia* harbours two genes, which code for this enzyme, the AOX1- and the AOX2-gene, whereat the AOX1-gene provides the by far greater portion of active alcohol oxidase in the cells. The expression of the AOX1-gene is regulated and induced by methanol. For the preferred expression system of this invention, the AOX1-gene was isolated and the AOX1-promotor was used for the expression of an arbitrary gene (Ellis et al., 1985; Koutz et al., 1989; Tschopp et al., 1987a).

The form of heterologous expression of the CEL I-enzyme in *Pichia pastoris* being preferred in this invention is the secretory form of protein expression. Secretory protein expression in *Pichia pastoris* has the advantage, that - because of the very low level of native protein secretion of this yeast - the major component of the total protein in the medium is constituted by the desired protein. This facilitates further steps of purification of the heterologous protein or even makes them potentially unnecessary.

The secretory mechanism preferably used in this expression method is based on the secretion signal α-factor of *Saccharomyces cerevisiae* (Barr et al., 1992), which is already integrated in the prefabricated expression vector pPIC 9 (Invitrogen).

A further reason for the preference of the *Pichia* systems e.g. to prokaryotic expression systems is the capability of *Pichia pastoris* to perform post-translational modification like e.g. the N-glycosidic affiliation of sugars, but without causing hyperglycosylation like it is e.g. the case with *S. cerevisiae* (Grinna and Tschopp, 1989; Tschopp et al., 1987b). Post-translational modifications can be crucial for the proper function of an enzyme.

The construct, which is preferred in this invention for the expression of the active CEL I-enzyme consists of the Cel I-6His-sequence-molecule, which is ligated in the appropriate orientation into the *Eco*RI restriction site of the expression vector pPIC 9, and which in this form has its open reading frame in fusion with the signal peptide. By means of cloning into the *Eco*RI restriction site of the vector, the CEL I-gene, preferably the Cel I-6His-construct, is put under the control of the AOX1-promotor being positioned 5'of the construct. For cloning purposes in *E.coli*, the vector provides both an ampicillin resistance and an *E.coli* origin of replication (FIG. 4).

It is necessary to integrate the gene into the yeast genome in order to achieve expression of the CEL I-gene. The integration, in the case being preferred herein, in accomplished by a homologous recombination, i.e. by a *crossing over* between the His4-locus on the chromosome and the His4-locus on the vector. The His4-gene of *Pichia pastoris* is used for the selection of stable transformants. For this purpose, the His4-gene, which is part of the histidine metabolism pathway, is present in the yeast genome in a mutated form, whereas it is present in the vector in the wildtype form.

For this reason, yeast cells without the integrated vector are not capable to grow on histidine-free medium, whereas yeast cells successfully transformed with pPIC9 are capable to form colonies on a medium containing no histidine. Since the vector lacks a yeast origin of replication, only yeast cell colonies can arise, in the founder cell of which a recombination has taken place between the plasmid and the yeast genome whereby the vector including the target gene has been integrated into the yeast genome.

The preferred technique of transformation in this method is the yeast transformation by means of electroporation. For this technique one adds 20-30 µg of linearised vector-DNA, purified by phenol extraction after linearisation, to 80 µl freshly competent cells of the yeast strain GS115 in a sterile cuvette. The CEL I-6His-pPIC9 construct in this method is linearized via the unique *SalI* restriction site (FIG. 4).

The electroporation was performed by means of a Gene Pulser II Systems (Biorad) employing 50 µF/200Ω/1,8V and a pulse time of about 10 msec.

Successfully transformed cells can be identified after an incubation period of 5 days on a histidine-free medium as properly grown colonies. Further evaluation in respect of a stable transformation was accomplished by a PCR-based detection of the target gene within the yeast genome.

As a starting primer for the PCR, we used a CEL I-specific primer Testf: 5'-ATGACCAGACTGTACTCCGTGTTC-3' (SEQ ID No.1; FIG. 2) and as an opposite primer, we used a primer being complementary to the vector cassette, the primer AOX3': 5'-GCAAATGGCATTCTGACATCC-3' (SEQ ID No.2; FIG. 4). The size of the expected PCR-product was about 1000 bp. As a positive control, the vector construct Cel I-6His-pPIC9 was used as a template. As negative controls, we used both genomic yeast DNA of a clone being transformed with the parental vector pPIC9 without a CEL I-insert, and a sample without a template in order to exclude "false positives" arising in consequence of contaminations.

In an example given herein for the preferred method, 16 clones of 20 yeast clones tested were identified as unambiguously positive (FIG. 5). For additional certainty in respect of the integration of the target gene into the yeast genome, PCR-positive clones were analysed by hybridising a respective Southern Blot with a CEL I-specific probe. Also here the controls used were a plasmid-DNA of the CEL I-6His-pPIC9-construct (positive control) and a genomic DNA being transformed with the parental vector pPIC9 without the CEL I-insert into yeast (negative control).

In one example given herein for the preferred method, a digoxigenin-labelled probe with a length of 262 bases was synthesised. The probe synthesis was accomplished by means of two oligonucleotides specifically annealing in the N-terminal region of the coding CEL I-

sequence, the oligonucleotides "Sonde f" ("probe f") 5'-ATGACCAGACTGTACTCCGTGTTC-3' (SEQ ID No. 3) and "Sonde r" ("probe r") 5'-GTCAGGGGTATCAATGAAATGTAA-3' (SEQ ID No.4; FIG. 2).

In one example given herein for the preferred method 10 clones of 12 clones being tested as PCR-positive were again confirmed to be positive by means of Southern hybridisation (FIG. 6). Clones being tested as undoubtedly positive in both test methods were used for expression.

The expression of the gene CEL I under the control of the AOX1-promotor (Ellis et al., 1985; Koutz et al., 1989; Tschopp et al., 1987a) was realised by the repressing/derepressing mechanism and subsequent induction, as described for *Saccharomyces cerevisiae* (Johnston, 1987). The exact employment in *Pichia pastoris* was realised according to the protocol of the *Pichia* expression system (Invitrogen).

Cultivation and growth of the clones was realised for two days in a medium containing glucose. Glucose acts as a repressor of the genes being controlled by the AOX1-promotor, thus blocking their transcription. After having reached a cellular density of $OD_{600}$ = about 3-5, expression was induced by changing the medium and adding the inductor methanol at a cellular density of $OD_{600}$ = 1. Expression was allowed for a period of 7-8 days with an addition of the metabolised carbon source methanol every 24 hours.

A further aspect of the invention refers to a recombinant, complete CEL I-protein produced by a method according to the invention.

Since the expressed protein is secreted in the preferred example for expressing CEL I, the desired enzyme was able to be easily purified from the supernatant of the expression culture by means of techniques known in the prior art. After having concentrated the proteins in the supernatant by a factor of about 200 by means of ultrafiltration tubes (Vivascience), the active CEL I-enzyme allowed to be used directly as a protein, which recognises and cleaves mismatch sequences.

Since *Pichia pastoris*, as already mentioned, secretes very low amounts of protein, the expressed and secreted CEL I-enzyme was ready for employment without further processing or purification steps.

A specific activity assay assessing functionality is required to test the enzyme's capability to recognise all of the eight base mismatch combinations.

In an exemplary manner, constructs were created for this purpose, which allowed to synthesise all of the eight mismatch combinations by means of the respective combinations of heterohybrids. For the application preferred in this method, the generation of these constructs was accomplished by the cloning of four oligonucleotides into the *Eco*RI/*Hind*III cleaved pUC19 vector. These oligonucleotides only differed at one single base position (FIG. 7). The four cloned fragments were able to be used as defined templates for the amplification of fragments using fluorescence-labelled oligonucleotides directly taking part in heterohybrid formation. According to the combination of amplification targets in hetero-hybrid synthesis, all of the eight mismatch combinations possible allowed to be generated.

The amplification of 237 bp fragments was accomplished by means of the fluorescence-labelled PUC19 F-primer 5'-FAM-GGATGTGCTGCAAGGCGAT-3' (SEQ ID No.5) and the fluorescence-labelled PUC19 R-primer 5'-JOE-GTGAGTTAGCTCACTCATTAG-3' (SEQ ID No.6). After heterohybrid formation of the partner fragments desired in the respective case was accomplished by a denaturation at 95°C for 10 min and subsequent gradual cooling, the activity assay was performed by incubating the heterohybrids with a 1:50 dilution of the CEL I-extract from the yeast expression supernatant at 47°C for 10 min.

CEL I cleaved one strand of the heterohybrids specifically at the site of the mismatch. After having applied the sample onto a GeneScan-gel, a 94 bp fragment and a 143 bp fragment were detected instead of the 237 bp fragments, as it was correspondingly the case for the opposite strand (FIG. 8).

The enzyme produced in this method by means of the artificially synthesised gene exactly displays its desired property, i.e. the precise recognition of all possible mismatch combinations as well as the subsequent incision into one strand at the phosphodiester bond immediately 3' of the detected base mismatch (Leykowski et al., 1998).

A further aspect thus relates to the use of the recombinantly produced CEL I-enzyme according to the invention for detecting both point mutations as well as larger mutated regions like e.g. deletions/insertions.

The appended sequence protocol shows:

SEQ ID No. 1: primer Testf: 5'-ATGACCAGACTGTACTCCGTGTTC-3',

SEQ ID No. 2: primer AOX3': 5'-GCAAATGGCATTCTGACATCC-3',

SEQ ID No. 3: Sonde f: 5'-ATGACCAGACTGTACTCCGTGTTC-3',

SEQ ID No. 4: Sonde r: 5'-GTCAGGGGTATCAATGAAATGTAA-3',

SEQ ID No. 5: F-primer: 5'-FAM-GGATGTGCTGCAAGGCGAT-3',

SEQ ID No. 6: fluorescence-labelled PUC19 R-primer:
5'-JOE-GTGAGTTAGCTCACTCATTAG-3',

SEQ ID No. 7: nucleic acid sequence of the mature CEL I-enzyme according to the invention after redraft for expressing the enzyme in yeast,

SEQ ID No. 8: amino acid sequence of the mature CEL I-enzyme, and

SEQ ID No. 9: presentation of the complete nucleotide sequence of the synthetic CEL I-gene.

The enclosed figures show:

FIG. 1: A depiction of the nucleic acid sequence (SEQ ID No. 7) required for encoding the mature CEL I-enzyme after redraft for expressing the enzyme in yeast. The amino acid sequence (SEQ ID No. 8) is also presented. Furthermore presented are base deviations from the published original sequence, which are shown in grey characters.

FIG. 2: A depiction of the complete nucleotide sequence of the synthetic CEL I-gene (SEQ ID No. 9). The sequence is given as a double strand, whereat the deoxyoligonucleotides necessary for synthesis are each printed in boldface on the respective strand. Moreover indicated are the sequence modifications like restriction sites, the Kozak sequence and the His-tag encoding sequences. The His-tag encoding sequence sections positioned at the N-terminus or the C-terminus are underlined; the underlined sequences were added either N-terminally or C-terminally, but not at both termini. The oligonucleotides Test f/Sonde f or Sonde r, which are required for several test experiments, are indicated by grey shading of the sequence.

FIG. 3: Schematic depiction concerning the synthesis of the artificial CEL I-gene by means of asymmetric PCR employing 16 overlapping deoxyoligonucleotides.

FIG. 4: Schematic depiction of the vector pPIC9 (Invitrogen), which is preferably used in this invention. The figure shows the integration of the artificial CEL I-gene (about 100 bp) into the *Eco*RI-restriction site of the vector; also shown is the oligonucleotide-primer AOX3' required for the PCR-test.

FIG. 5: PCR-result for the verification of the integration of the CEL I-gene into the genome of several yeast clones. Genomic DNA as a template was isolated from 20 yeast clones to be tested. Genomic DNA of two non-transformed yeast clones served as a template for the negative control (-). Purified vector-DNA of two original constructs served as a template for the positive control. The blank sample comprised water in order to exclude contaminations.

16 clones of 20 clones to be tested are unambiguously positive. Corresponding to the positive controls, they show a band of about 1000 bp. Negative controls and blank sample are free of this signal. As a molecular weight marker (M) the "1 kb Plus DNA Ladder" (Gibco) is shown.

FIG. 6: Result of the Southern hybridisation for the further verification of 12 yeast clones tested before by the PCR method. Both as a positive control and as a size control, the 262 bp CEL I-specific probe was hybridised to plasmid-DNA (+) (Construct shown in FIG. 4).

As a negative control, the probe was hybridized to genomic yeast DNA of a clone containing the parental vector pPIC 9 without the CEL I-insert (-).

FIG. 7: Construct for generating defined heterohybrids for performing a specific activity assay of the CEL I-enzyme. The two depicted synthetic oligonucleotides are constructed such, that they can be directly ligated into the *Eco*RI/*Hind*III digested pUC19-vector after annealing (what is possible due to their complementary nature). Each of the two deoxyoligonucleotides is present in fourfold version. The letters Y and Z each symbolise all of the four possible bases. In consequence, all of the eight possible base mismatches (AA/TT/CC/GG/AC/AG/TC/TG) can be synthesised depending on the combination of the oligonucleotides.

FIG. 8: Exemplary result of the specific activity assay of the recombinant CEL I-enzyme in case of the base mismatch AA. The 137 bp PCR-product containing the mismatch is specifically incised at the mutation. Partly the Fam-labelled strand and partly the Joe-labelled strand is cleaved. Since the mutation is not exactly located in the middle of the PCR-product, the fragment is asymmetrically cleaved into a 94 bp fragment and a 143 bp fragment.

**References:**

Barbon J J, McKenzie M and Cotton R G H: Mutation detection using fluorescent enzyme mismatch cleavage with T4 endonuclease VII. Electrophoresis 20: 1162-1170 (1999)

Barr K A, Hopkins S A and Sreekrishna K: Protocol for efficient secretion of HAS developed from Pichia pastoris. Pharm Eng 12: 48-51 (1992)

Ellis S B, Brust P F, Koutz P J, Waters A F, Harpold M M and Gingeras T R: Isolation of alcohol oxidasse and two other methanol regulatable genes from yeast, *Pichia pastoris*. Mol Cell Biol 5: 1111-1121 (1985)

Golz S, Birkenbihl RP and Kemper B: Improved large-scale preparation of phage T4 endonuklease VII overexpressed in E.coli. DNA Res; 2 (6): 277-284 (1995)

Grinna L S and Tschopp J F: Size distribution and general structural features of N-linked oligosaccarides from the methylotrophic yeast *Pichia pastoris*. Yeast 5: 107-115 (1989)

Johnston M: A model fungal gene regulatory mechanism: the GAL genes of *saccharomyces cerevisiae*. Microbiol Rev 51: 458-476 (1987)

Koutz P J, Davis G R, Stillman C, Barringer K, Cregg J M and Thill G: Structural comparison of the *Pichia pastoris* alcohol oxidase genes. Yeast 5: 167-177 (1989)

Kowalski D and Sandford J R: Action of mung bean nuclease on supercoiled PM2 DNA. J Biol Chem 257: 7820-7825 (1982)

Kozak M: An analysis of 5'-noncoding sequence from 699 vertebrate messengers RNAs. Nuclei Acid Res 15: 8125-8148 (1987)

Kozak M: Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl AcadSci USA 87: 8301-8305 (1990)

Loeb L A and Silber J R: S1 nuklease does not cleave DNA at single-base mis-matches. Biochim Biophys Acta 656 (2): 256-264 (1981)

Norberg T, Klaar S, Lindqvist L, Lindahl T, Ahlgren J and Bergh J: Enzymatic mutation detection method evaluated for detection of *p*53 mutations in cDNA from breast cancers. Clinical Chemistry 47 (5): 821-828 (2001)

Oleykowski C A, Brinson Mullis C R, Chang D W and Yeung A T: Incision at nucleotide insertions/deletions and base pair mismatches by the SP nuclease of spinach. Biochemistry 38: 2200-2205 (1999)

Oleykowski C A, Bronson Mullins C R, Godwin A K and Yeung A T: Mutation detection using a novel plant endonuclease. Nucleic Acids Research 26 (20): 4597-4602 (1998)

Outchkourov NS, Stiekema WJ, Jongsma MA. Optimization of the Expression of Equistatin in Pichia pastoris. Protein Expr Purif 2002 Feb;24(1):18-24

Paula de Mattos Areas A, Leonor Sarno de Oliveira M, Raul Romero Ramos C, Sbrogio-Almeida M, Raw I, Ho P. Synthesis of cholera toxin B subunit gene: cloning and expression of a functional 6XHis-tagged protein in Escherichia coli. Protein Expr Purif 2002 Aug;25(3):481

Solaro P C, Birkenkamp K, Pfeiffer P and Kemper B: Endonuklease VII of phage T4 triggers mismatch correction in vitro. L Mol Biol 230 (3): 868-877 (1993)

Tschopp J F, Brust P F, Cregg J M, Stillmann C and Gingeras T R: Expression of the *lacZ* gene from two methanol regulated promoters in *Pichia pastoris*. Nucleic Acids Res 15: 3859-3876 (1987a)

Tschopp J F, Sverlow G, Kosson R, Craig W and Grinna L: High level secretion of glycosylated invertase in the methylotrophic yeast *Pichia pastoris*. Bio Technology 5: 1305-1308 (1987b)

Yang B, Wen X, Kodali N S, Oleykowski C A, Miller C G, Kulinski J, Besack D, Yeung J A, Kowalski D and Yeung A T: Purifiction, cloning and characterization of the CEL I nuclease. Biochemistry 39:3533-3541 (2000)

## Claims:

1. Method for producing a nucleic acid sequence, which codes for the complete CEL I-protein and allows to be recombinantly expressed in host cells, comprising the steps of:
   - Providing the sequence coding for the CEL I-protein from a suitable organism, in particular from *Apium graveolens* L., and
   - Adequately modifying the codon frequency of the sequence to be expressed in comparison to the native sequence, whereat this modification is performed with regard to the host organism to be used for expression, by means of
   a) Partition of the planned sequence into an even number x, especially 8, overlapping regions,
   b) Synthesis of 2x mutated oligonucleotides, in particular oligonucleotides 1-16 to 16-16, each of which comprise the entire length of one overlapping region of both strands of the coding sequence,
   c) First PCR-amplification in order to produce x/2, in particular 4, overlapping fragments under employment of the oligonucleotides of step b)
   d) Second PCR-amplification in order to produce x/4, in particular 2, overlapping regions under employment of the fragments of step c)
   e) Third PCR-amplification in order to produce x/8, in particular 1, fragment, which comprises the coding region of CEL I.

2. Method according to claim 1, characterised in that instead of step e) the following steps are performed:
   e') Cloning the fragments generated in step d) into a suitable vector, and
   f) Appropriately digesting the vectors and ligating the fragments in order to fuse said fragments to form the complete fragment comprising the coding region of CEL I.

3. Method according to claim 1 or 2, characterised in that the codon frequency of the sequence to be expressed is modified according to the codon frequency of yeast.

4. Method according to any one of the claims 1 to 3, characterised in that furthermore nucleotides are attached that encode for additional N-terminal or C-terminal amino acid tags, in particular tags being comprised of 6 histidines.

5. Method according to any one of the claims 1 to 4, characterised in that the oligonucleotides synthesised in step (c) have an average length of 70 nucleotides and in each case overlap at about 20 bases.

6. Method for producing a recombinant, complete CEL I-protein from *Apium graveolens* L.,
comprising
- performing the method according to any one of the claims 1 to 5, and
- expressing the nucleic acid sequence by means of a suitable expression system.

7. Method according to claim 6, characterised in that a vector is used as an expression system, whereat this vector is selected from the pPIC 9, pPIC 3, 5 and pQE-vectors.

8. Method according to claim 6 or 7, characterised in that the nucleic acid sequence is expressed in a host cell selected from *Hansenula polymorpha*, *Pichia pastoris*, *Saccharomyces cerevisiae*, HeLa-cells, CHO-cells, *Toxoplasma gondii* and *Leishmania*.

9. Method according to claim 8, characterised in that the employed *Pichia pastoris* strain is the strain GS115.

10. Recombinant, complete CEL I-protein from *Apium graveolens* L., produced according to any one of the claims 6 to 9.

11. Complete DNA-sequence of the CEL I-protein suitable for the expression of the CEL I-protein or expressible parts of this sequence derived from *Apium graveolens*, obtainable by a method according to any one of the claims 1 to 5.

12. DNA-sequence of the CEL I-protein of *Apium graveolens* according to claim 11, characterised in that furthermore nucleotides are attached that encode for additional N-terminal or C-terminal amino acid tags, in particular tags being comprised of 6 histidines.

13. DNA-sequence of the CEL I-protein of *Apium graveolens* according to claim 11 or 12, characterised in that the sequence at its both ends provides restriction sites for restriction endonucleases, which are absent in the remaining sequence and in a vector to be employed.

14. Complete DNA-sequence of the CEL I-protein suitable for the expression of the CEL I-protein of *Apium graveolens* according to SEQ ID No. 7 or parts of this sequence.

15. Host organism expressing a DNA-sequence according to any one of the claims 11 to 14.

16. Host organism according to claim 15, selected from *Hansenula polymorpha*, *Pichia pastoris*, *Saccharomyces cerevisiae*, HeLa-cells, CHO-cells, *Toxoplasma gondii* and *Leishmania*.

17. Host organism according to claim 16, characterised in that the employed *Pichia pastoris* strain is the strain GS115.

18. Use of the recombinantly produced CEL I-enzyme according to claim 10 for detecting point mutations as well as larger mutated regions like e.g. deletions/insertions.

Fig.:1

```
                                                M  T  R  L  Y  S  V  F  F
                                                ATGACCAGACTGTACTCCGTGTTCTT
                                                   GC  T  A  T  T

 L  L  L  A  L  V  V  E  P  G  V  R  A  W  S  K  E  G  H  V  M
TCTACTTCTCGCCCTTGTCGTGGAGCCCGGTGTAAGAGCTTGGTCAAAGGAAGGACATGTTA
  TT GT G  T     A  T  A  G     T     C   AGC  A     C     C

 T  C  Q  I  A  Q  D  L  L  E  P  E  A  A  H  A  V  K  M  L
TGACCTGTCAGATTGCCCAGGACCTTCTTGAGCCAGAAGCCGCTCATGCGGTAAAGATGTTG
   A     A     G     T  GT G           A  A     T           C

 L  P  D  Y  A  N  G  N  L  S  S  L  C  V  W  P  D  Q  I  R  H
TTGCCTGATTATGCCAACGGAAACTTATCAAGCCTATGTGTTTGGCCAGATCAGATCCGTCA
  A  G  C     T  T  C        GTCG  G     G     T     A  T  A

 W  Y  K  Y  R  W  T  S  S  L  H  F  I  D  T  P  D  Q  A  C  S
TTGGTACAAGTACCGTTGGACCTCCTCCTTACATTTCATTGATACCCCTGACCAAGCATGTT
C            A  G     TAG   TC C        C     A     T     C

  F  D  Y  Q  R  D  C  H  D  P  H  G  G  K  D  M  C  V  A  G
CCTTTGACTATCAACGTGACTGTCATGATCCCCATGGTGGGAAGGACATGTGCGTTGCCGGC
 A     T  C  GA A                 A     A                 T     T  A

 A  I  Q  N  F  T  S  Q  L  G  H  F  R  H  G  F  S  D  R  R  Y
GCGATTCAAAATTTCACCTCTCAATTGGGGCATTTCCGTCACGGTACAAGTGATAGGCGATA
  C              A     GC T  A        C  T  A   TC    C  T

 N  M  T  E  A  L  L  F  L  S  H  F  M  G  D  I  H  Q  P  M  H
CAATATGACTGAAGCTTTGCTCTTCCTTTCACACTTCATGGGAGACATTCATCAACCTATGC
T        A  G     T  A  TT A  C              T        G

  V  G  F  T  S  D  M  G  G  N  S  I  D  L  R  W  F  R  H  K
ATGTGGGATTTACTTCCGACATGGGCGGTAATAGTATTGATTTGAGGTGGTTTCGTCATAAA
    T        AAGT  T     A  A  C     A      C  C        C  C

 S  N  L  H  H  V  W  D  R  E  I  I  L  T  A  A  A  D  Y  H  G
TCAAACCTGCATCACGTCTGGGATCGAGAGATCATTCTAACTGCTGCTGCTGATTATCACGG
  C        C  T  T     A        T     T  A     A  A     C  T

 K  D  M  H  S  L  L  Q  D  I  Q  R  N  F  T  E  G  S  W  L  Q
AAAGGATATGCATTCCTTGCTTCAAGACATTCAGAGAAATTTTACGGAGGGTTCTTGGTTGC
T           C  TC C  A        A     G  C     A      AG

  D  V  E  S  W  K  E  C  D  D  I  S  T  C  A  N  K  Y  A  K
AAGACGTGGAATCTTGGAAAGAATGCGATGATATCTCTACTTGTGCAAACAAATATGCTAAG
    T  T     C     G     T                  C  C  T  G

 E  S  I  K  L  A  C  N  W  G  Y  K  D  V  E  S  G  E  T  L  S
GAGTCAATTAAGTTGGCTTGTAACTGGGGGTATAAGGATGTAGAAAGTGGAGAGACATTGTC
   AGT  A  AC A  C           T  C  A     T   TC   C  A  TC

 D  K  Y  F  N  T  R  M  P  I  V  M  K  R  I  A  Q  G  G  I  R
GGATAAATATTTTAACACGCGAATGCCAATTGTTATGAAACGTATCGCCCAAGGAGGAATCA
A        C  C     AA              C        G  A  T  G  T     C

  L  S  M  I  L  N  R  V  L  G  S  S  A  D  H  S  L  A
GATTAAGCATGATTCTTAACCGTGTCCTGGGTTCGTCTGCTGACCATTCGTTGGCATAATAA
 T   TC       T  G     A  T  T  AAGC  C  A  T     T
```

Fig.:2

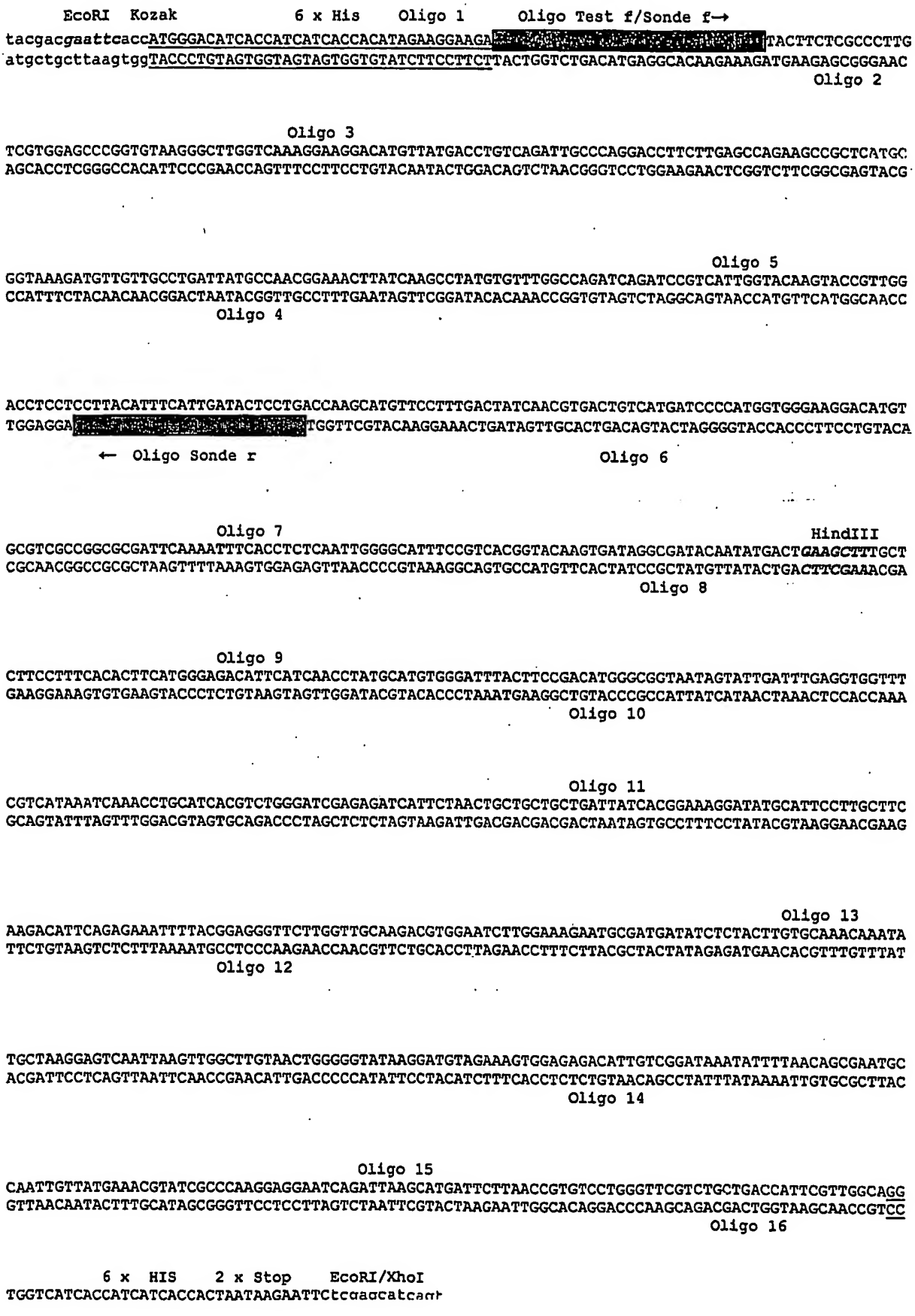

EcoRI Kozak
6 x His
Oligo 1
Oligo Test f/Sonde f→
tacgacgaattcaccATGGGACATCACCATCATCACCACATAGAAGGAAGA
TACTTCTCGCCCTTG
atgctgcttaagtggTACCCTGTAGTGGTAGTAGTGGTGTATCTTCCTTCTTACTGGTCTGACATGAGGCACAAGAAAGATGAAGAGCGGGAAC
Oligo 2
Oligo 3
TCGTGGAGCCCGGTGTAAGGGCTTGGTCAAAGGAAGGACATGTTATGACCTGTCAGATTGCCCAGGACCTTCTTGAGCCAGAAGCCGCTCATGC
AGCACCTCGGGCCACATTCCCGAACCAGTTTCCTTCCTGTACAATACTGGACAGTCTAACGGGTCCTGGAAGAACTCGGTCTTCGGCGAGTACG
Oligo 5
GGTAAAGATGTTGTTGCCTGATTATGCCAACGGAAACTTATCAAGCCTATGTGTTTGGCCAGATCAGATCCGTCATTGGTACAAGTACCGTTGG
CCATTTCTACAACAACGGACTAATACGGTTGCCTTTGAATAGTTCGGATACACAAACCGGTGTAGTCTAGGCAGTAACCATGTTCATGGCAACC
Oligo 4
ACCTCCTCCTTACATTTCATTGATACTCCTGACCAAGCATGTTCCTTTGACTATCAACGTGACTGTCATGATCCCCATGGTGGGAAGGACATGT
TGGAGGA
TGGTTCGTACAAGGAAACTGATAGTTGCACTGACAGTACTAGGGGTACCACCCTTCCTGTACA
← Oligo Sonde r
Oligo 6
Oligo 7
HindIII
GCGTCGCCGGCGCGATTCAAAATTTCACCTCTCAATTGGGGCATTTCCGTCACGGTACAAGTGATAGGCGATACAATATGACTGAAGCTTTGCT
CGCAACGGCCGCGCTAAGTTTTAAAGTGGAGAGTTAACCCCGTAAAGGCAGTGCCATGTTCACTATCCGCTATGTTATACTGACTTCGAAACGA
Oligo 8
Oligo 9
CTTCCTTTCACACTTCATGGGAGACATTCATCAACCTATGCATGTGGGATTTACTTCCGACATGGGCGGTAATAGTATTGATTTGAGGTGGTTT
GAAGGAAAGTGTGAAGTACCCTCTGTAAGTAGTTGGATACGTACACCCTAAATGAAGGCTGTACCCGCCATTATCATAACTAAACTCCACCAAA
Oligo 10
Oligo 11
CGTCATAAATCAAACCTGCATCACGTCTGGGATCGAGAGATCATTCTAACTGCTGCTGCTGATTATCACGGAAAGGATATGCATTCCTTGCTTC
GCAGTATTTAGTTTGGACGTAGTGCAGACCCTAGCTCTCTAGTAAGATTGACGACGACGACTAATAGTGCCTTTCCTATACGTAAGGAACGAAG
Oligo 13
AAGACATTCAGAGAAATTTTACGGAGGGTTCTTGGTTGCAAGACGTGGAATCTTGGAAAGAATGCGATGATATCTCTACTTGTGCAAACAAATA
TTCTGTAAGTCTCTTTAAAATGCCTCCCAAGAACCAACGTTCTGCACCTTAGAACCTTTCTTACGCTACTATAGAGATGAACACGTTTGTTTAT
Oligo 12
TGCTAAGGAGTCAATTAAGTTGGCTTGTAACTGGGGGTATAAGGATGTAGAAAGTGGAGAGACATTGTCGGATAAATATTTTAACAGCGAATGC
ACGATTCCTCAGTTAATTCAACCGAACATTGACCCCCATATTCCTACATCTTTCACCTCTCTGTAACAGCCTATTTATAAAATTGTGCGCTTAC
Oligo 14
Oligo 15
CAATTGTTATGAAACGTATCGCCCAAGGAGGAATCAGATTAAGCATGATTCTTAACCGTGTCCTGGGTTCGTCTGCTGACCATTCGTTGGCAGG
GTTAACAATACTTTGCATAGCGGGTTCCTCCTTAGTCTAATTCGTACTAAGAATTGGCACAGGACCCAAGCAGACGACTGGTAAGCAACCGTCC
Oligo 16
6 x HIS
2 x Stop
EcoRI/XhoI
TGGTCATCACCATCATCACCACTAATAAGAATTCtcgagcatcgagt

Fig.: 3

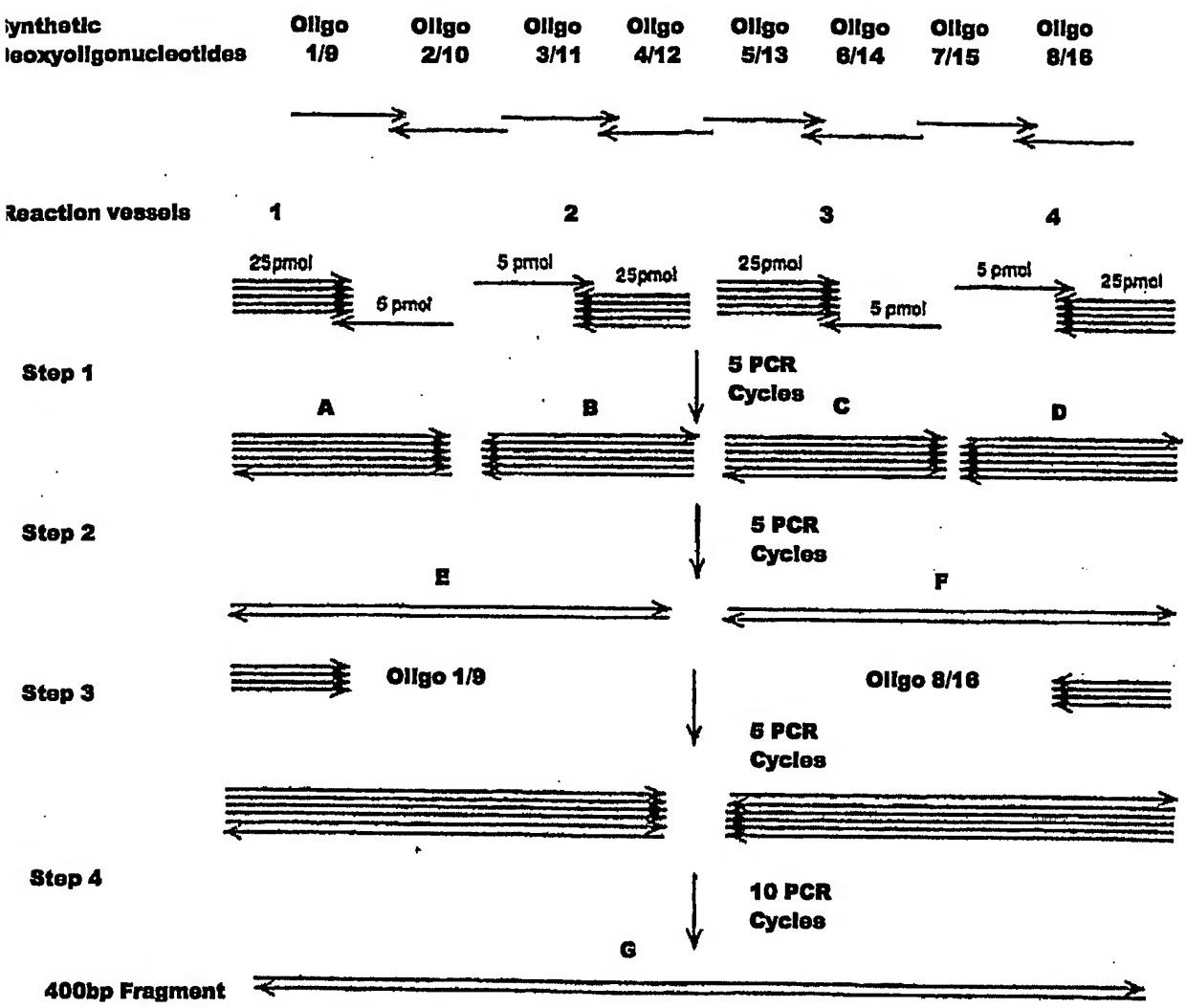

ynthetic
leoxyoligonucleotides
Oligo 1/9
Oligo 2/10
Oligo 3/11
Oligo 4/12
Oligo 5/13
Oligo 6/14
Oligo 7/15
Oligo 8/16
Reaction vessels
1
2
3
4
25pmol
6 pmol
5 pmol
25pmol
25pmol
5 pmol
5 pmol
25pmol
Step 1
5 PCR Cycles
A
B
C
D
Step 2
5 PCR Cycles
E
F
Step 3
Oligo 1/9
Oligo 8/16
5 PCR Cycles
Step 4
10 PCR Cycles
G
400bp Fragment

Fig.:4

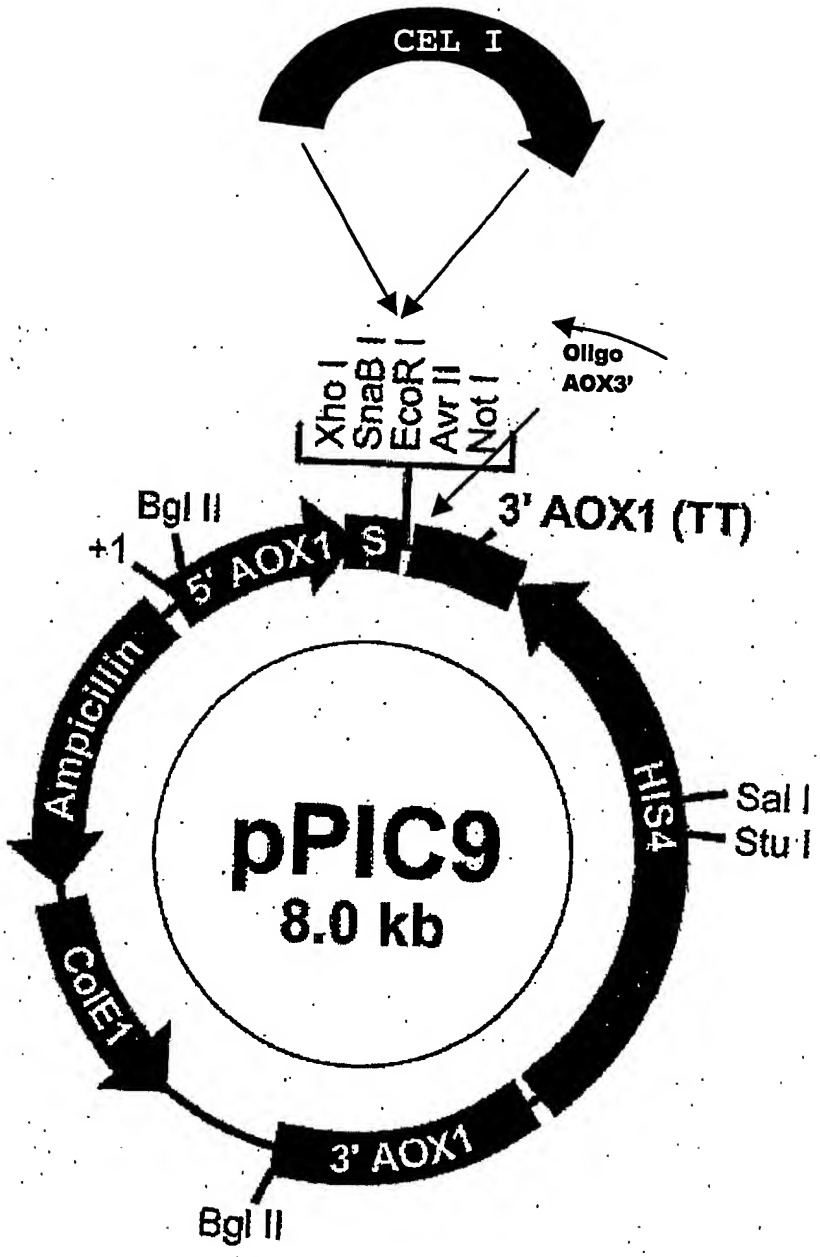

CEL I
Xho I
SnaB I
EcoR I
Avr II
Not I
Oligo
AOX3'
Bgl II
+1
5' AOX1
S
3' AOX1 (TT)
Ampicillin
pPIC9
8.0 kb
HIS4
Sal I
Stu I
ColE1
3' AOX1
Bgl II

Fig.:5

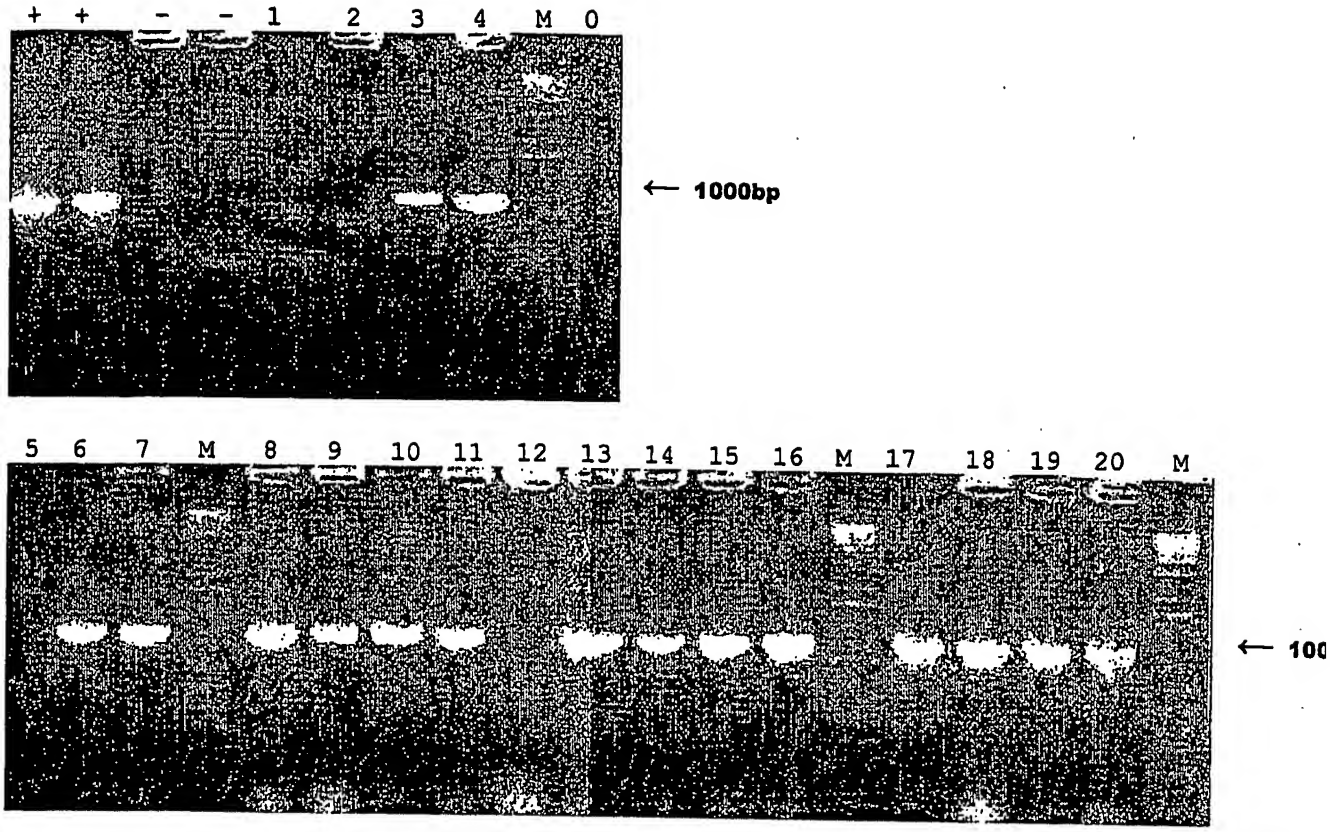


Fig.:6

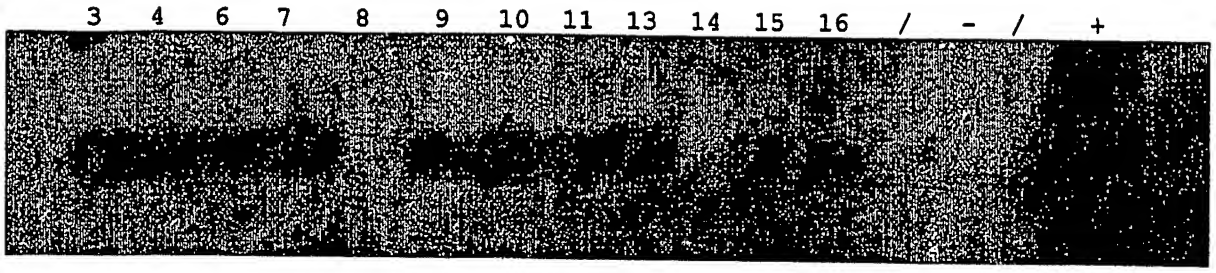


Fig.:7

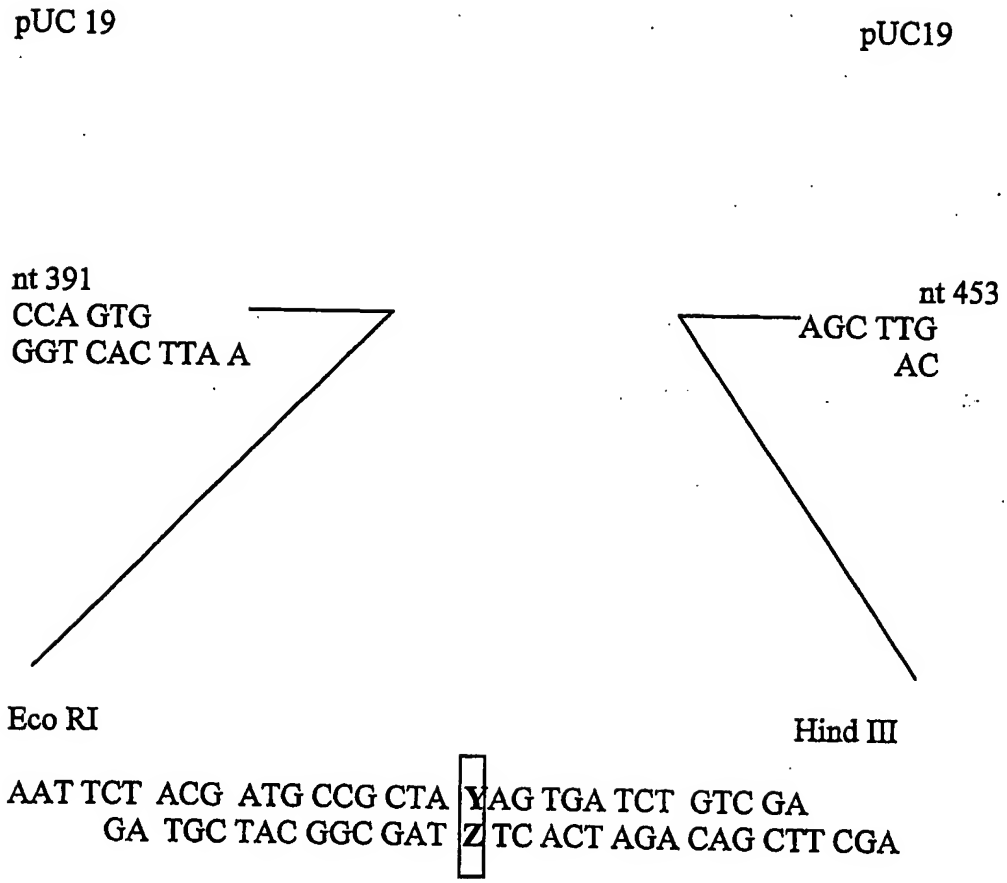
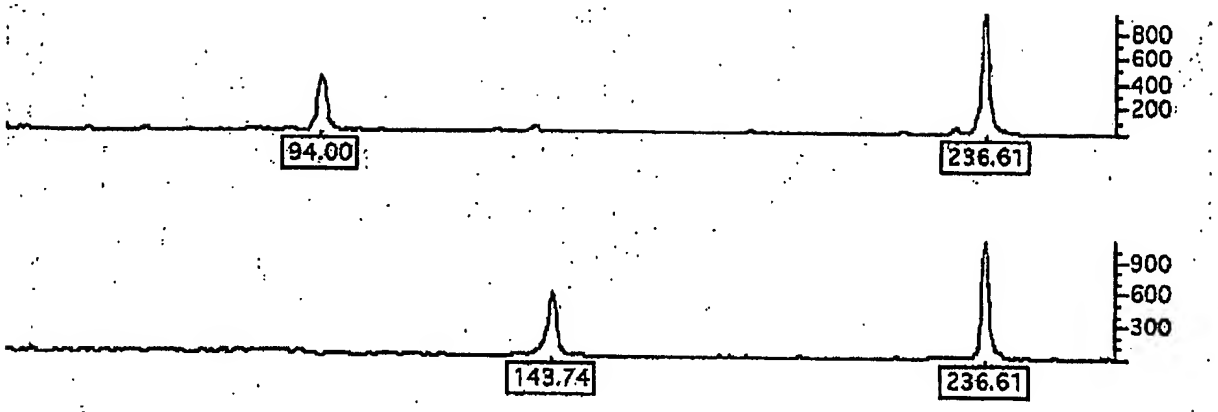

pUC 19
pUC19
nt 391
CCA GTG
GGT CAC TTA A
nt 453
AGC TTG
AC
Eco RI
Hind III
AAT TCT ACG ATG CCG CTA
GA TGC TAC GGC GAT
Y
Z
AG TGA TCT GTC GA
TC ACT AGA CAG CTT CGA

Fig.:8



SEQUENCE LISTING

<110> Biopsytec Analytik GmbH

<120> Mutated nucleic acid of a CelI-endonuclease and methods for producing the recombinant full-length CelI-protein

<130> B30098PCT

<160> 9

<170> PatentIn version 3.2

<210> 1
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 1
atgaccagac tgtactccgt gttc 24

<210> 2
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 2
gcaaatggca ttctgacatc c 21

<210> 3
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 3
atgaccagac tgtactccgt gttc 24

<210> 4
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 4

<210> 5
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 5
ggatgtgctg caaggcgat 19

<210> 6
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 6
gtgagttagc tcactcatta g 21

<210> 7
<211> 895
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 7
atgacgcgat tatattctgt gttctttctt ttgttggctc ttgtagttga accgggtgtt 60

agagcctgga gcaaagaagg ccatgtcatg acatgtcaaa ttgcgcagga tcttgttgga 120

gccagaagca gcacatgctg taaagatgct gttaccggac tatgctaatg gcaacttatc 180

gtcgctgtgt gtgtggcctg atcaaattcg acactggtac aagtaccgtt ggacctcctc 240

cttacatttc attgataccc ctgaccaagc atgttccttt gactatcaac gtgactgtca 300

tgatccccat ggtgggaagg acatgtgcgt tgccggcgcg attcaaaatt tcacctctca 360

attggggcat ttccgtcacg gtacaagtga taggcgatac aatatgactg aagctttgct 420

cttcctttca cacttcatgg gagacattca tcaacctatg catgtgggat ttacttccga 480

catgggcggt aatagtattg atttgaggtg gtttcgtcat aaatcaaacc tgcatcacgt 540

ctgggatcga gagatcattc taactgctgc tgctgattat cacggaaagg atatgcattc 600

cttgcttcaa gacattcaga gaaattttac ggagggttct tggttgcaag acgtggaatc 660

ttggaaagaa tgcgatgata tctctacttg tgcaaacaaa tatgctaagg agtcaattaa 720

ttttaacacg cgaatgccaa ttgttatgaa acgtatcgcc caaggaggaa tcagattaag 840

catgattctt aaccgtgtcc tgggttcgtc tgctgaccat tcgttggcat aataa 895

<210> 8
<211> 296
<212> PRT
<213> Apium graveolens

<400> 8

Met Thr Arg Leu Tyr Ser Val Phe Phe Leu Leu Leu Ala Leu Val Val
1 5 10 15

Glu Pro Gly Val Arg Ala Trp Ser Lys Glu Gly His Val Met Thr Cys
20 25 30

Gln Ile Ala Gln Asp Leu Leu Glu Pro Glu Ala Ala His Ala Val Lys
35 40 45

Met Leu Leu Pro Asp Tyr Ala Asn Gly Asn Leu Ser Ser Leu Cys Val
50 55 60

Trp Pro Asp Gln Ile Arg His Trp Tyr Lys Tyr Arg Trp Thr Ser Ser
65 70 75 80

Leu His Phe Ile Asp Thr Pro Asp Gln Ala Cys Ser Phe Asp Tyr Gln
85 90 95

Arg Asp Cys His Asp Pro His Gly Gly Lys Asp Met Cys Val Ala Gly
100 105 110

Ala Ile Gln Asn Phe Thr Ser Gln Leu Gly His Phe Arg His Gly Phe
115 120 125

Ser Asp Arg Arg Tyr Asn Met Thr Glu Ala Leu Leu Phe Leu Ser His
130 135 140

Phe Met Gly Asp Ile His Gln Pro Met His Val Gly Phe Thr Ser Asp
145 150 155 160

Met Gly Gly Asn Ser Ile Asp Leu Arg Trp Phe Arg His Lys Ser Asn
165 170 175

Leu His His Val Trp Asp Arg Glu Ile Ile Leu Thr Ala Ala Ala Asp
180 185 190

Tyr His Gly Lys Asp Met His Ser Leu Leu Gln Asp Ile Gln Arg Asn

Phe Thr Glu Gly Ser Trp Leu Gln Asp Val Glu Ser Trp Lys Glu Cys
210 215 220

Asp Asp Ile Ser Thr Cys Ala Asn Lys Tyr Ala Lys Glu Ser Ile Lys
225 230 235 240

Leu Ala Cys Asn Trp Gly Tyr Lys Asp Val Glu Ser Gly Glu Thr Leu
245 250 255

Ser Asp Lys Tyr Phe Asn Thr Arg Met Pro Ile Val Met Lys Arg Ile
260 265 270

Ala Gln Gly Gly Ile Arg Leu Ser Met Ile Leu Asn Arg Val Leu Gly
275 280 285

Ser Ser Ala Asp His Ser Leu Ala
290 295

<210> 9
<211> 986
<212> DNA
<213> Artificial

<220>
<223> Initial sequence from Apium graveolens, codon usage fitted to Pichia pastoris

<400> 9
tacgacgaat tcaccatggg acatcaccat catcaccaca tagaaggaag aatgaccaga 60

ctgtactccg tgttctttct acttctcgcc cttgtcgtgg agcccggtgt aagggcttgg 120

tcaaaggaag gacatgttat gacctgtcag attgcccagg accttcttga gccagaagcc 180

gctcatgcgg taaagatgtt gttgcctgat tatgccaacg gaaacttatc aagcctatgt 240

gtttggccag atcagatccg tcattggtac aagtaccgtt ggacctcctc cttacatttc 300

attgatactc ctgaccaagc atgttccttt gactatcaac gtgactgtca tgatccccat 360

ggtgggaagg acatgtgcgt cgccggcgcg attcaaaatt tcacctctca attggggcat 420

ttccgtcacg gtacaagtga taggcgatac aatatgactg aagctttgct cttcctttca 480

cacttcatgg gagacattca tcaacctatg catgtgggat ttacttccga catgggcggt 540

aatagtattg atttgaggtg gtttcgtcat aaatcaaacc tgcatcacgt ctgggatcga 600

gagatcattc taactgctgc tgctgattat cacggaaagg atatgcattc cttgcttcaa 660

gacattcaga gaaattttac ggagggttct tggttgcaag acgtggaatc ttggaaagaa 720

tgcgatgata tctctacttg tgcaaacaaa tatgctaagg agtcaattaa gttggcttgt 780

gaatgccaat tgttatgaaa cgtatcgccc aaggaggaat cagattaagc atgattctta 900

accgtgtcct gggttcgtct gctgaccatt cgttggcagg tggtcatcac catcatcacc 960

actaataaga attctcgagc atcagt 986

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 03/11210

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N9/22 C12N15/55 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | YANG BING ET AL: "Purification, cloning, and characterization of the CEL I nuclease" BIOCHEMISTRY, AMERICAN CHEMICAL SOCIETY. EASTON, PA, US, vol. 39, no. 13, 4 April 2000 (2000-04-04), pages 3533-3541, XP002222862 ISSN: 0006-2960 page 3535, left-hand column; figure 3 --- -/-- | 10,11, 14,15,18 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents:

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 2 February 2004 | 17/02/2004 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016 | Bucka, A |

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 03/11210

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | OLEYKOWSKI C A ET AL: "MUTATION DETECTION USING A NOVEL PLANT ENDONUCLEASE" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 26, no. 20, 1998, pages 4597-4602, XP002943289 ISSN: 0305-1048 the whole document | 10,18 |
| X | WO 01 62974 A (FOX CHASE CANCER CT ;YEUNG ANTHONY T (US)) 30 August 2001 (2001-08-30) figure 2; examples 8-10 | 10,18 |
| Y | page 32 -page 34 | 1-9, 11-17 |
| Y | WOO JUNG HEE ET AL: "Gene optimization is necessary to express a bivalent anti-human anti-T cell immunotoxin in Pichia pastoris" PROTEIN EXPRESSION AND PURIFICATION, vol. 25, no. 2, July 2002 (2002-07), pages 270-282, XP002266044 ISSN: 1046-5928 page 271, right-hand column page 280; table 2 | 1-9, 11-17 |
| Y | WO 01 66693 A (NOVOZYMES AS) 13 September 2001 (2001-09-13) the whole document | 1-9, 11-17 |
| A | OUTCHKOUROV NIKOLAY S ET AL: "Optimization of the expression of equistatin in Pichia pastoris" PROTEIN EXPRESSION AND PURIFICATION, vol. 24, no. 1, February 2002 (2002-02), pages 18-24, XP002266045 ISSN: 1046-5928 cited in the application the whole document | 1-18 |
| A | SINCLAIR GRAHAM ET AL: "Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, Pichia pastoris." PROTEIN EXPRESSION AND PURIFICATION, vol. 26, no. 1, October 2002 (2002-10), pages 96-105, XP002266046 ISSN: 1046-5928 the whole document | 1-18 |
| A | WO 97 46701 A (FOX CHASE CANCER CENTER) 11 December 1997 (1997-12-11) the whole document | 1-18 |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 03/11210

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 0162974 A | 30-08-2001 | US 6391557 B1 | 21-05-2002 |
| | | AU 4721201 A | 03-09-2001 |
| | | CA 2400441 A1 | 30-08-2001 |
| | | EP 1257671 A1 | 20-11-2002 |
| | | WO 0162974 A1 | 30-08-2001 |
| WO 0166693 A | 13-09-2001 | AU 3920501 A | 17-09-2001 |
| | | WO 0166693 A1 | 13-09-2001 |
| | | US 2003022280 A1 | 30-01-2003 |
| WO 9746701 A | 11-12-1997 | US 5869245 A | 09-02-1999 |
| | | AU 721493 B2 | 06-07-2000 |
| | | AU 3284597 A | 05-01-1998 |
| | | CA 2256526 A1 | 11-12-1997 |
| | | EP 0964929 A1 | 22-12-1999 |
| | | JP 2000511774 T | 12-09-2000 |
| | | WO 9746701 A1 | 11-12-1997 |